Capstone Project on Data Warehouse Design and ETL Implementation Using Informatica

Objective:
The goal of this capstone project is to design and implement a data warehouse by extracting insights from the provided datasets, leveraging Informatica as the primary ETL tool. Participants will gain hands-on experience in data modeling, ETL processes, and SQL querying to derive meaningful business insights.

Project Scope:

1. **Data Understanding and Modeling:**
   ○ Analyze the provided datasets to identify dimensions and facts.
   ○ Design a data model comprising dimension and fact tables.
2. **ETL Implementation:**
   ○ Define a data loading strategy for both dimensions and fact tables.
   ○ Create and populate staging tables using Informatica.
   ○ Develop mappings and workflows to load dimension and fact tables in adherence to the defined strategy.
3. **Data Analysis:**
   ○ Write SQL queries using the dimension and fact tables to derive actionable insights and validate data accuracy.

Best Practices to Follow:

● **Avoid Hardcoding:**
  Ensure no hardcoding at the workflow level. Use parameters for connections and database names.
● **Workflow Dependencies:**
  Create workflow dependencies using event_wait, touch commands, or PMCMD commands.
● **Naming Conventions:**
  Use proper naming conventions for tables, mappings, transformations, sessions, tasks, and workflows for clarity and consistency.

Expected Outcomes:

● A fully functional data warehouse comprising well-designed dimension and fact tables.
● Successful data extraction, transformation, and loading (ETL) using Informatica.
● SQL-based insights derived from the data model.
● Demonstrated adherence to coding and workflow management best practices.

Project Technical Stack:

Technology	Tool Used	Version
Data Integration Tool	Informatica	v10.2
Database	SQL Server	2022
Scheduler	Informatica Inbuilt Scheduler	v10.2

Capstone Project Evaluation Criteria:

Evaluation Component	Weightage (Marks)
Code Walkthrough	25
Data Warehouse Design	25
Reporting/Analytical Presentation	25
Presentation	25

Total Marks: 100

Please find the list of project datasets assigned in the following table:

Full Name	Project Assigned	Datasets
Aishi Biswas	Employees Payroll in Los Angeles	https://drive.google.com/file/d/1qz7uFGjsBppnpkVR1kMAEgHNd6JytNMf/view
Kruthik Ballari	U.S._Chronic_Disease_Indicators	https://drive.google.com/file/d/10OQL1OiN6z-UmLLrB5GKI7AmbaPEqBSo/view
Challa Vyshnavi Reddy	U.S. Counties Covid 19 dataset	https://drive.google.com/file/d/1pU-vM6INGsRn5gMqoA7r-fXn0TFwvMga/view
Ishita Jindal	Government Finance Statistics Dataset	https://drive.google.com/file/d/16v8Us3MeNQ6Tw6TlH363RFFg2IozNkEN/view
Ganapathineedi Deekshita	Loan Approval Dataset	https://drive.google.com/file/d/1VV7de4bgbUjd4BpiCfkGUqe0NClpoxef/view
Ishwarya Devanathan	Andhra Pradesh Health Dataset	https://drive.google.com/file/d/1JHkqOGGDkh2b78WAIlvMad9Jck4N0e0s/view
Kolla Sahithi	Crop Production Dataset	In the attachment
Mirudhula L	Super Store Dataset	In the attachment
Nainala Bhagya Vinaya Sindhu	Ethnic Group Census Dataset	In the attachment

Muralidharan V	Co2 Emissions Dataset	In the attachment
Sagar Jamuar	US General-Election Dataset	In the attachment
Shreyas Patil	School Enrollment Dataset	In the attachment
Virajit	Stocks Trader dataset	In the attachment
Yadlapati Sindhu Priya	World Exports dataset	In the attachment